

Monitoring Markov Chains

Mahyar Karimi

Institute of Science and Technology Austria

September 13, 2023



Coins

- | Given two (*maybe* fair) coins:



- | How different are the coins?

Coins

- | Given two (*maybe* fair) coins:

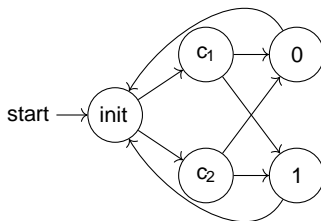


- | How different are the coins?

$$P(1 | c_1) \quad P(1 | c_2)$$

Coins in Markov Chains

- | Let's turn this into a Markov chain:

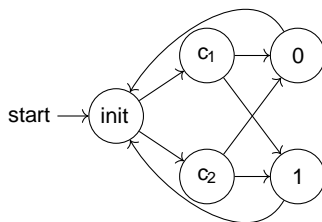


- | Observing this Markov chain Stream of states:

init ; c₁ ; 1 ; init ; c₂ ; 0 ; init ; c₁ ; 0 ; :::

Coins in Markov Chains

- | Let's turn this into a Markov chain:



- | Observing this Markov chain Stream of states:

init ; c₁ ; 1 ; init ; c₂ ; 0 ; init ; c₁ ; 0 ; :::

- | How to estimate $P(1 | c_1)$ $P(1 | c_2)$?

What We Developed

- I Techniques for estimating expressions over P_{ij}
 - Frequentist and Bayesian: two well-known paradigms in statistics

What We Developed

- | Techniques for estimating expressions over $\mathcal{P}(j)$
 - Frequentist and Bayesian: two well-known paradigms in statistics
- | => Monitors for transitions in Markov chains

What We Developed

- | Techniques for estimating expressions over P_{ij}
 - Frequentist and Bayesian: two well-known paradigms in statistics
- | => Monitors for transitions in Markov chains
- | => Monitoring fairness properties!

The Frequentist, The Bayesian

I The frequentist

- Probabilities Long-run frequencies
- Estimating the ground truth.

The Frequentist, The Bayesian

I The frequentist

- Probabilities Long-run frequencies
- Estimating the ground truth.

I The Bayesian

- Markov chains sampled from a given distribution (or belief).
- No ground truth! Only prior belief.

Frequentist Monitor

| Monitor: Input stream! Inteval stream:

A : ! I₁ ;1

¹Probably Approximately Correct

Frequentist Monitor

| Monitor: Input stream! Interval stream:

$$A : \quad ! \quad I_1 ; 1$$

| The monitor gives PAC¹-style guarantees (for given):

$$P \left[\sum A \right] \leq \epsilon$$

¹Probably Approximately Correct

Frequentist Monitor

- | Monitor: Input stream! Interval stream:

$$A : \quad ! \quad I_1 ; 1$$

- | The monitor gives PAC¹-style guarantees (for given):

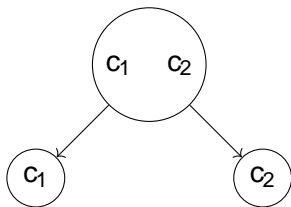
$$P \uparrow 2 A (\times) \quad 1$$

- | Intervals (= Bounding with Hoeffding's inequality.

¹Probably Approximately Correct

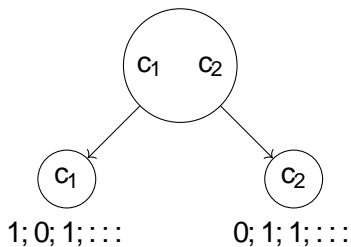
Frequentist and Coins

I For $P(1 | c_1)$ $P(1 | c_2)$:



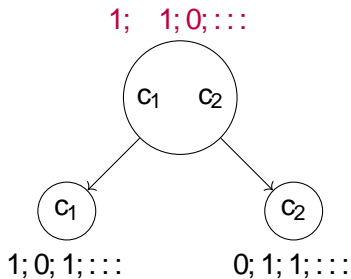
Frequentist and Coins

I For $P(1 | c_1)$ $P(1 | c_2)$:



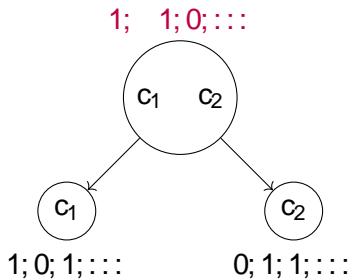
Frequentist and Coins

I For $P(1 | c_1)$ $P(1 | c_2)$:



Frequentist and Coins

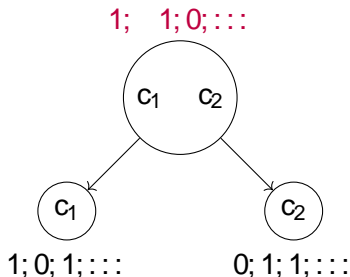
- For $P(1 | c_1)$ $P(1 | c_2)$:



- New RV with the same expected value as the subtraction

Frequentist and Coins

- For $P(1 | c_1)$ $P(1 | c_2)$:



- New RV with the same expected value as the subtraction
- A few other techniques for multiplication and division

The Bayesian

- | Fixed prior distribution² over Markov chains (prior belief)
- | Obtaining posterior belief after observing x

²Matrix beta distribution

Bayesian Monitor

- | The monitor guarantees the following (for given ϵ):

$$P(\|M - A\| \geq \epsilon) \leq \epsilon$$

Bayesian Monitor

- The monitor guarantees the following (for given ϵ):

$$P(\|M_j(x) - A(x)\| \geq \epsilon) \leq \epsilon$$

- Core component: efficient, incremental computation of $E_j(x)$:

$$E_j(x) = E_{j-1}(x) \frac{c_{ab}(x) + d_{ab}}{c_{ab}(x)} + \frac{c_a(x)}{c_a(x) + d_a}$$

Base case comes from

Bayesian Monitor

- The monitor guarantees the following (for given ϵ):

$$8\epsilon^2 : P \left(\sum_{j=1}^n (M_j - A_j) \geq \epsilon \right) \leq 1$$

- Core component: efficient, incremental computation of $E \left(\sum_{j=1}^n (M_j - A_j) \right)$:

$$E \left(\sum_{j=1}^n (M_j - A_j) \right) = E \left(\sum_{j=1}^n (M_j - A_j) \right) \frac{c_{ab}(\epsilon) + d_{ab}}{c_{ab}(\epsilon)} = \frac{c_a(\epsilon)}{c_a(\epsilon) + d_a}$$

Base case comes from

- Intervals (ϵ) = Bounding with Chebyshev's inequality.

Monitoring Regular Expressions?

| So far: $P(b^j a)$

Monitoring Regular Expressions?

| So far: $P(b \ j \ a)$

| $P(b_1 :: b_k \ j \ a)$ is not that different:

$$P(b_1 :: b_k \ j \ a) = P(b_1 \ j \ a) \cdot P(b_k \ j \ b_{k-1})$$

| ! Monitoring *star-free* regular expression r : $P(r \ j \ a)$.